PRODUCT SHEET: 3' mRNA-SEQ ON SINGLE CELLS

The GenomEast platform provides a 3 'mRNA-seq service on single cells by combining 10X Genomics micro-fluidic technology and Illumina high-throughput sequencing. RNA-seq on single cells (scRNA-seq) aims to analyze the transcriptome of the different cells in a given sample. It thus makes it possible to study cell heterogeneity and to characterize, on the basis of their gene expression profile, the different subpopulations of cells present in a sample.

1 10X Genomics 3'mRNA-seq technology

The Chromium Controller developed by 10X Genomics is an equipment which allows, from a cell suspension, to individually encapsulate from 500 to 10,000 cells in droplets, containing all the necessary reagents for cell lysis and cDNA synthesis as well as a gel bead containing specific oligonucleotides (Fig.1). These oligonucleotides consist of:

- a PCR primer (Truseq Read 1 or Nextera Read 1);
- a barcode (BC), unique to each bead and therefore to each cell;
- a Unique Molecular Identifier (UMI), a short random sequence unique to each oligonucleotide on the bead and therefore to each transcript or molecule captured;
- a poly (dT) tail for the capture of polyA + RNA from target cells or specific capture sequences (Capture Seq 1 et 2) for the simultaneous analysis of surface proteins.
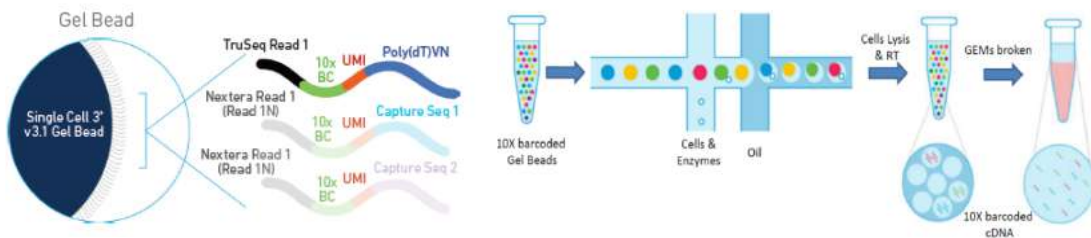

**Fig.1: Principle of technology (*source: 10X Genomics)*

After isolation in the nano droplets, the cells are lysed by heating and the mRNAs thus released are retained on the beads by complementary hybridization thanks to their polyA tail. A reverse transcription reaction converts the mRNAs into cDNAs labeled with specific barcodes and UMI. The emulsion is then broken, the hydrogel beads are dissolved and the sequencing library is finalized on all of the barcoded cDNAs by adding sequencing adapters and PCR amplification.

On the platform, the final libraries (Fig. 2) are sequenced on a NextSeq 2000. Read 1 containing the cell barcode and UMI is used to demultiplex the cells during the primary analysis. Read 2 is used to identify and quantify transcripts.
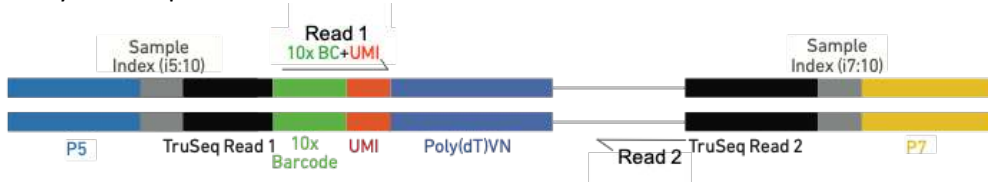

**Fig.2: Structure of the final 3'mRNA-seq libraries (*source: 10X Genomics)*

Only the preparation of 3'mRNA-seq libraries is provided as standard service on our platform. Additional applications such as CITE-seq or CellPlex can nevertheless be carried out in the form of scientific collaboration.

2   Planning of experiments

All service requests must be submitted on the platform's web interface at the following address: http://ngs-lims.igbmc.fr/ by creating a new request after opening a user account.

When submitting his request, the project manager must specify in the project description: the total number of samples (i.e. number of different cell suspensions) to be processed simultaneously on the Chromium, the total number of cells he wishes to load per well and the total number of cells he expects to capture. He must also assess the biological risk linked to his cell samples and inform the platform of the appropriate containment measures (Bio Safety Level BSL1 ou BSL2). **The platform will not support any project involving samples identified as BSL3 or BSL4.**

As far as possible, the project manager must plan the date of the experience with the platform at least 2 weeks before the run. In addition, in order to be able to complete the experience in the usual working hours, it is desirable to deposit the cells on the platform from Monday to Thursday at the latest at 4 p.m.

The capture rate on Chromium is estimated at around 65% (i.e. 65% of the droplets contain a cell). The multiplet rate depends on the total number of cells loaded on the Chromium. It is advisable to refer to Table 1 below to determine the optimal number of cells to be loaded according to the tolerated multiplet rate and the desired final number of captured cells.

Multiplet Rate (%)	# of Cells Loaded	# of Cells Recovered
~0.4%	~800	~500
~0.8%	~1,600	~1,000
~1.6%	~3,200	~2,000
~2.3%	~4,800	~3,000
~3.1%	~6,400	~4,000
~3.9%	~8,000	~5,000
~4.6%	~9,600	~6,000
~5.4%	~11,200	~7,000
~6.1%	~12,800	~8,000
~6.9%	~14,400	~9,000
~7.6%	~16,000	~10,000

**Tab.1: Multiplet rate and number of captured cells as a function of the number of loaded cells (*source: 10X Genomics)*

3   Services provided
1. Starting samples checking (optional but recommended):
   - Cell counting and viability test with trypan blue.

GenomEast Platform
CERBM GIE – IGBMC
1 rue Laurent Fries
67404 Illkirch Cedex
France

+33 3 88 65 34 26
thibault@igbmc.fr
jost@igbmc.fr
genomeast.igbmc.fr

2. Library preparation:
   - Preparation of libraries according to 10X Genomics recommendations using "Chromium Single Cell 3' Reagent kits";
   - Quantification and verification of the quality of final libraries by capillary electrophoresis (Agilent Fragment Analyzer or Bioanalyzer 2100).

3. Sequencing using Illumina NextSeq 2000 technology:
   - Paired-end sequencing of 28 bases for read 1 and 85 bases for read 2;
   - Note that the 10X Genomics libraries are sequenced in the presence of ~2% of Illumina PhIX control library to counter the lack of diversity in the read 1 bases.

4. Primary data analysis:
   - Demultiplexing and generation of FASTQ files;
   - Adapter dimer removal;
   - Sequences quality check;
   - Generation of a report summarizing the methods used in the primary data analysis pipeline as well as the results obtained.

5. Downstream data analysis (optional, see section 7 for more information).

4   Sample preparation by project manager

According to 10X Genomics recommendations, **under optimal conditions**, the cell suspension loaded on the Chromium must have the following characteristics:
- Concentration : 700 à 1000 cells/µl;
- Resuspension buffer : PBS + 0.04 % BSA;
- Viability rate : >70%;
- Cell size : 5 to 30 µm.

The total volume of cell suspension loaded on the machine should be ≤41.3 µl. The minimum cell concentration required to achieve a capture rate of ~5000 cells per sample is then 200 cells/µl.

To increase the chances of achieving the desired cell recovery rate, it is imperative to avoid cell aggregates and to eliminate cell debris and dead cells as much as possible before loading onto the Chromium Controller. These factors will indeed have a negative impact on the recovery rate since dead cells and debris will be captured in the same way as living cells.

Different protocols for cell dissociation and preparation of cell suspensions are available at these two addresses:
- https://support.10xgenomics.com/single-cell-gene-expression/sample-prep
- http://www.worthington-biochem.com/tissuedissociation/default.html

When he comes to deposit his cells on the platform, the project manager must provide the following information:
- the concentration of its cells;
- their viability rate;

GenomEast Platform
CERBM GIE – IGBMC
1 rue Laurent Fries
67404 Illkirch Cedex
France

+33 3 88 65 34 26
thibault@igbmc.fr
jost@igbmc.fr
genomeast.igbmc.fr

- the total volume of cell suspension.

5   Quality controls performed by the platform

The ultimate success of the experiment is closely linked to the accuracy of cell counting and the determination of their viability. It is also important to minimize the time between preparing the suspension and loading it onto the Chromium Controller.

If the viability rate is much less than 70%, in particular if the total number of cells is small (<1000 cells), the platform reserves the right to stop the experiment and cancel the run on the Chromium.

If desired, the project leader can ask the platform to use its own cell count results for loading onto the Chromium Controller. In these conditions, the platform will carry out a verification count only after loading onto the machine if the starting material allows it.

1. Checking of starting cell suspension (optional but recommended)	
Total number of cells to load on Chromium Controller (N) (Hemacytometer)	$1000 \leq N \geq 10000$
Total cell volume to load on Chromium Controller (V)	$5\ \mu l \leq V \geq 41.3\ \mu l$
% Cell viability (Bleu trypan)	$\geq 70\%$
2. Library preparation	
Library profile (capillary electrophoresis)	Average size ranging from 200 to 600 bp
Library purity (capillary electrophoresis)	Limited presence of adapter dimers (120-130 pb band)
3. Sequencing and primary data analysis	
Total number of clusters* per project (including PhiX)	≥ Total number of clusters specified in the "Requested services" section from the submission form (pdf file that can be downloaded from the LIMS http://ngs-lims.igbmc.fr, in the "Document" tab for each project)
Quality Score (Phred score) > 30	$\geq 85\%$ of bases

Number of reads in single-read and number of reads ÷ 2 in paired-end

6   Results delivery

For each sample, the following results are made available to the project manager:
- Raw sequencing data (nucleotide sequences in FASTQ format. The files contain reads passing quality filters and do not contain dimer adapter sequences).

In addition to these sample files, two files are provided for each project:
- A report (PDF file) presenting the number of raw reads, the percentage of bases with a Phred quality score greater than 30 and the size of raw files (FASTQ) to download;
  A text file providing the MD5 string of each FASTQ file to download The project manager is responsible for downloading his files, checking their integrity from MD5 strings and storing them. A documentation is available on the following webpage: http://genomeast.igbmc.fr/wiki/doku.php?id=help:md5.

A data delivery email informs the project manager that he can download his sequence data using a login and password on the platform's FTP server.

GenomEast Platform
CERBM GIE – IGBMC
1 rue Laurent Fries
67404 Illkirch Cedex
France

+33 3 88 65 34 26
thibault@igbmc.fr
jost@igbmc.fr
genomeast.igbmc.fr

According to the "GenomEast Platform terms and conditions of business", the project manager is responsible for his data to be saved and archived on its own. Following their transfer to the Beneficiary, the Platform guarantees the conservation of raw data only for a limited period of six months.

7 Downstream analysis (optional)

Data analysis is not part of the standard service but can be done in collaboration between the project manager and the platform. The following analyses can be performed:

- Extraction of barcodes and UMIs from reads 1, alignment to a reference genome of reads 2, filtering of barcodes and UMIs, generation of a gene x barcode matrix, cell filtering;
- Analysis of gene expression in all cells: dimension reduction, clustering, differential expression analysis;
- Single-cell trajectory inference.

The above list is not exhaustive and we recommend that project managers who would like to collaborate with the platform for data analysis to contact us before starting their project so that we can help them define the analyses that best fit their needs.